

**<http://hpg.genopole.fr/>**

Constitution d'une base de données pour servir à l'histoire du projet  
génomique

Rapport de fin de contrats, K. Gay (CNRS) & J-F Picard (IRH CNRS), juillet 2013

## **Sommaire**

1 - PRESENTATION DE L'OPERATION 'BDD HPG'

2 - CONTENU DE LA BASE

- Archives
- Publications
- Entretiens

3 - FONCTIONNEMENT DE LA BASE

- Fonctionnement
- Structure
- Modalités d'accès

## **1 - PRESENTATION DE L'OPERATION 'BDD HPG'**

Le 19 juin 2009, le programme 'histoire du projet génome' (HPG) est lancé par Jim Watson au Cold Spring Harbor laboratory. Jean Weissenbach accepte la responsabilité de mener la recherche sur la participation française au 'Human Genome Program'.

En septembre 2009 à la suite de la réunion au CSHL, un Comité pour l'histoire du projet génome (HPG) est constitué au Génopole d'Evry avec H. Cann (CEPH), J. Weissenbach (Genoscope), J-F Prud'Homme (Genopole), P. Tambourin (dir. du Génopole) et D. Guthleben (CNRS).

Le 3 février 2010 se tient une première réunion du 'Comité HPG' au CNRS en présence de J. Weissenbach, J-F. Prud'homme, D. Guthleben, L. Tiennot Herment, M. Karkatchian et A. Bullo. La marche à suivre est fixée pour collecter la documentation nécessaire à la réalisation d'une base de données (BDD) à partir des archives de l'AFM-Généthon, du CEPH, du Génopole, CNS et du CNRS. A. Bullo est la correspondante pour les archives du Genethon, E. Guibert pour l'AFM et M-C Bachelier pour les archives du CNRS.

Le 1er juillet 2010 J-F Picard est contractualisé par Genopole d'Evry en vue de sa participation à la collecte des matériaux nécessaires à la constitution de la BDD (une convention précédente AFM-Génopole-CNRS du 6 juillet 2001 concernant l'histoire de la génomique' lui ayant donné l'opportunité de réunir une abondante documentation).

Le 15 septembre 2010, un 'Contrat collaboration' destiné à officialiser le programme HPG est passé entre le Genopole, l'AFM et le CNRS (Comité pour l'histoire du CNRS, M. Blay). Dans son préambule, il est précisé que le comité de pilotage du programme HPG est présidé par J. Weissenbach et son secrétariat assuré par J - F Prud'Homme. Le Génopole, l'AFM et le CNRS mettent leurs archives à la disposition des chercheurs engagés dans le programme, le Génopole se chargeant d'héberger la BDD destinée au stockage des documents recueillis .

En novembre 2010, en relation avec le CERMES3 où il participe à des recherches portant sur l'organisation de la recherche médicale, J-F Picard propose d'infléchir le programme HPG vers les relations entre le programme génome humain et la médecine.

Le 30 mai 2011, lors d'une réunion du Comité, E. Guiraud, mandatée par la présidente de l'AFM, demande de dresser un bilan d'étape nécessaire à la reconduction de la participation de l'Association au programme HPG. A défaut de quoi celle-ci déciderait de se retirer d'un programme marqué par certains dysfonctionnements énumérés ci dessous :

- Du fait d'un cahier des charges mal conçu, la BDD initiale a révélé des défauts de fonctionnement (absence d'indexation lisible des documents, pas de thésaurus, lourdeur de manipulation)

- A défaut des procédures adéquates, la collation d'archives scientifiques (cahiers de laboratoires, données numériques) s'est révélée irréalisable dans les délais impartis.

- Certains fonds d'archives n'ont pu être inventoriés faute des autorisations d'accès nécessaires (cas du CEPH)

Le 14 septembre 2011, D. Guthleben transmet au Génomole un rapport signifiant la fin de la participation du CNRS au programme HPG (il annonce la préparation d'un livre avec l'AFM et le lancement avec Michel Morange [ENS Ulm] d'un séminaire 'histoire de séquençage' ).

Le 20 septembre 2011, lors d'une réunion du programme Génomole, J-F Picard propose de refondre la base de données 'histoire du projet génome' avec l'aide d'une assistante (K. Gay, CNRS-CERMES3), le chargement de la première BDD étant interrompu depuis le mois de mai précédent. L'objectif est de fondre les divers fonds documentaires constitués depuis 2010, comme la documentation en cours de collecte dans la poursuite du programme HPG, en vue de leur incrémentation dans une BDD entièrement refondue qui inclurait :

- le fonds réuni et numérisé dans la première BDD par O. Le Faou et D Guthleben (AFM, CNRS)
- les documents réunis et scannés par J-F Prud'homme à partir des archives du CNS-Génomole (papiers J Weissenbach) et de l'analyse de la presse scientifique
- les documents réunis et numérisés par J-F Picard (prog. 'Histoire de la génomique', arch. Inserm, papiers H. Cann, A. Danchin, B. Jordan et J. Haiech) , etc.

Le 3 janvier 2012 une réunion organisée par P. Tambourin au Génomole avec J-F Prud'homme et J-F Picard recadre le programme HPG et donne lieu à deux contrats passés entre le Génomole d'Evry, J-F Picard et K Gay.

Le 12 juillet 2012, du fait des lacunes de la base en matière d'archives purement scientifiques, la quatorzième (et dernière réunion) du programme HPG entérine la non participation de l'IFRIS (P B Joly, A Kaufmann) à ce programme de recherche et confirme les modalités d'élaboration de la BDD telle qu'elle est présentée dans le présent rapport.

## 2 - CONTENU DE LA BASE

La base de données représente un fonds numérisé de plus de 4200 documents intéressant l'histoire du projet génome et certains de ses prolongements, en particulier ceux ayant trait à la recherche médicale. Selon le cahier des charges initial, le centre de gravité de cette BDD est constitué par la documentation collectée de manière systématique entre 1982 (fusion UMF-AMF et mise en place du CEPH) et 2003 (conclusion du programme de séquençage du génome humain). Mais il est apparu judicieux d'y incrémenter un lot de données se reportant aux périodes encadrantes. C'est ainsi que ce fonds a été alimentée par une série de documents (archives et publications) concernant les prémises du projet génome, issus des années 1940 à 1980 (développement de la génétique moléculaire et des biotechnologies) et d'une sélection de documents postérieurs à 2003 destinés à illustrer certains prolongements du programme génome, voire à documenter le déroulement du programme 'histoire du projet génome'. Si ce fonds d'archives numérisées ne peut prétendre à l'exhaustivité, de par son ampleur il constitue une source essentielle pour documenter la genèse, le déroulement et certains aboutissements des programmes génomes, particulièrement en France. Les documents réunis proviennent des fonds d'archives répertoriés ci-dessous, d'un important lot de publications et d'une série d'entretiens réalisés dans le cadre de différents programmes de recherche.

### ARCHIVES

Archives du Centre d'étude du polymorphisme humain (CEPH) : papiers et carnets de labo Howard Cann, Jean-Marc Lalouël. 'Newsletters', correspondance et notes de services, publications diverses, etc. (le fonds d'archives du CEPH n'a pu être inventorié)

Archives de l'Association française contre les myopathies (AFM) : procès verbaux du Conseil d'administration et de son Bureau, du conseil scientifique, appels d'offres, budgets, etc. Correspondance et notes de Bernard Barataud, Téléthon, etc.

AFM-Généthon : PV du Directoire et du Conseil scientifique, rapports d'activité et de prospective, Genexpress. Généthon II et III (programme GVPN), Appels d'offres, etc. Notes et correspondance de Bernard Barataud, relations avec Genset, Sanofi, etc.

Archives du CNRS : correspondance et notes issues du dept. SDV, nomenclature des labos et appels d'offres dans le domaine de la génomique, archives du GIP Génome (GREG), PV. du Conseil d'administration et 'Lettre du GREG', etc. Bioinformatique, programme européen de séquençage du génome de la levure, séquençage de B. Subtilis, etc.

Archives du Centre national de séquençage CNS-Génoscope et du Centre national de génotypage (CNG). PV des Comités de pilotage et rapports d'activité et de prospective, correspondance de J. Weissenbach avec Eurogem, Wellcome Trust (Sanger Center, Genetics and Medicine Historical Network), Human Genome Prog. US (NHGRI, newsletter 'Human Genome Project Information'), HUGO (arch. CORDIS), etc.

Archives de l'Inserm : rapports d'activité, appels d'offres et rapports d'activité de prospective du conseil scientifique. Dossiers de laboratoires (J Dausset, J Frezal, C. Griscelli et A Fischer, M Lamy, J. Feingold, etc.)

Les documents conservés par Bertrand Jordan et Jacques Haiech concernent l'action du ministère de la Recherche (lancement du GIP Génome, GREG, programme Très Grand Séquençage (TGS), Consortium national de recherche en génomique, Genopole d'Evry, les papiers de A. Danchin le séquençage de B. Subtilis et la bioinformatique, ceux de Daniel Pardo le CIT-Med. et le Génopole PACA.

Quelques cahiers de laboratoire ont pu être réunis : H. Cann (période CEPH), J-M Lalouel (d°) et F Galibert (période Sanger)

L'ensemble des notes et fiches documentaires issues du programme 'Histoire de la génomique' (Génopole-CNRS, N. Givernaud, J F Picard 2002) sont évidemment inclus dans la BDD.

## PUBLICATIONS

Ce fonds documentaire comporte un important lot d'articles issus de la littérature scientifique portant les programmes génomes, humains et autres, en particulier 'Nature', 'Science' (dépouillement quasi systématique) 'PNAS', 'CRAS', 'La Recherche', 'Médecine/Science', 'Biofutur', 'Bull. Ac. Med.', etc. comme de certaines revues de vulgarisation scientifique.

En ce qui concerne la presse d'information, un dépouillement du journal 'Le Monde' a été effectué de manière systématique dans les archives électroniques pour la période 1987-2010.

Enfin la BDD inclut des notes de lecture ou certains chapitres d'ouvrages (extraits) consacrés à la description du programme génome humain. De même on y a inséré en version intégrale certains travaux universitaires (thèses et mémoires) réalisés sur le sujet.

## ENTRETIENS

La transcription des entretiens réalisés avec les acteurs témoins dans le cadre du programme 'Histoire de la génomique' (2002) et ceux effectuées dans le cadre du programme HPG ou du séminaire 'histoire de séquençage' ont été versés dans la BDD (documents de travail et textes validés par les intéressés). Il convient de

signaler le fait que le Génopole ne détient pas le copyright des textes déjà publiés dans le cadre d'autres programmes de recherche (CNRS, Inserm ou autres).

Charles Auffray (Généthon)  
Bernard Barataud (AFM)  
André Boué (Inserm)  
Serge Braun (AFM)  
René Cadoret (d°)  
Michel Callon (CSI ENM)  
Howard Cann (CEPH)  
Marie Anne Chancerel (CNRS)  
Daniel Cohen (CEPH)  
Michel Cohen-Solal (Inserm)  
Antoine Danchin (GREG)  
Jean Dausset (CEPH)  
Jacques Demaille (CNRS)  
Dominique Duguet (I. de myologie)  
Jean-Marc Egly (IGBMC)  
Michel Fardeau (Inserm)  
Josué Feingold (d°)  
Alain Fischer (CHU Necker)  
Francis Galibert (CNRS)  
André Goffeau (U Louvain)  
Claude Griscelli (Inserm)  
François Gros (I. Pasteur)  
Jacques Haiech (Min Rec.)  
Jacques Hanoune (Inserm)  
Bertrand Jordan (CIML)  
Axel Kahn (ICGM)  
Jean-Claude Kaplan (CHU Cochin)  
Philippe Kourilsky (I. Pasteur)  
Jean-Marc Lalouël (HHMI)  
François Leterrier (AFM)  
Bernard Malissen (CIML)  
Robert Manaranche (Genethon)  
Jean-Louis Mandel (IGBMC)  
Claude Mawas (CIML)  
Eric Molinié (AFM)  
Arnold Munnich (CHU Necker)  
Daniel Pardo (Genopole PACA)  
Jean-François Prud'homme (Généthon)  
Frédéric Révah (Généthon)  
Jean Rosa (CHU Mondor)  
Piotr Slonimski (CGM CNRS)  
Pierre Tambourin (Genopole)  
Bruno Tocqué (ExonHit)  
Jean Weissenbach (CNS Genoscope)  
Raymond L. White (U. Utah)

### 3 - FONCTIONNEMENT ET STRUCTURE DE LA BASE DE DONNEES

#### FONCTIONNEMENT

Le fonds documentaire ainsi constitué est chargé dans une BDD relationnelle sous SQL, installée sur un site dynamique hébergé sur le serveur informatique du Génopole d'Evry, URL <http://hpg.genopole.fr/>. Le chargement de la base a été effectué par des rédacteurs dotés d'un code utilisateur spécifique fourni par le webmestre du Génopole d'Evry, Alain Astgen. Ces documents ont été indexés dans la base de données sous forme de notices, via une interface web développée sous le système de gestion de contenu open-source "Drupal". Ainsi, ce fonds documentaire forme le contenu d'un site internet dynamique permettant son exploitation et son éventuel développement selon les dispositions qui seront arrêtées par le Génopole d'Evry.

#### STRUCTURE

La BDD est interrogeable en langage naturel. Chaque document y est répertorié dans une notice identifié par :

- Un numéro à huit chiffres (**année/mois/quantième**) renvoyant à sa date de production (i. e. classement chronologique).
  
- Son **Titre** (ou le cas échéant son objet)
  
- Par une série de **Mots-clés** : Actions incitatives, Actions thématiques programmées (ATP), ADN, ADN recombinant, AFM, Anthropologie, ARN, Associations, Bacillus subtilis, Banques de gènes, Base de données, Biochimie, Biodiversité, Bioinformatique, Biologie cellulaire, Biologie moléculaire, Biomédecine, Biotechnologies, Biothérapie, Budget, Cancer, Carte d'identité génétique, Carte génétique, Carte physique, Cartes, Cartes intégrées, Cartographie du génome, cDNA, Cellules souches, Centre national de génotypage (CNG), Centre national de séquençage (CNS), CEPH, Chromosome (1, 8, 10, 11, 13, 14, 15, 16, 18, 20, 21, 22), Chromosomes artificiels, Clonage, Clonage positionnel, CNRS, Code génétique, Cytogénétique, Dépistage, Dépistage anténatal, Diabète, Diagnostic génétique, Diagnostic préimplantatoire, Diagnostic prénatal, Economie, Enseignement, Enzymes de restriction, Epidémiologie génétique, Epigénétique, Escherichia coli, Essais thérapeutiques, Ethique, Ethnologie, Eugénisme, Evolution, Familles, Fondation Jean Dausset, Genatlas, Gène de détermination sexuelle, Gène recombinant, Gènes, Genethon (2, 3), Genethon-Industries, Génétique, Génétique de la levure, Génétique des populations, Génétique des tumeurs, Génétique des virus, Génétique du développement, Génétique fonctionnelle, Génétique humaine, Génétique inverse, Génétique médicale, Génétique moléculaire, Genexpress, Génie génétique, Génome humain, Génomes, Génomique, Genopole, Génoscope, Génotypage, Genset, Glaucome, Greg, Groupes sanguins, Hématologie,



Hémophilie, Hérité, Histoire, Histoire du programme génome, Histoire du programme génome humain, HLA, Homosexualité, Hypertension artérielle, Imabio, Imagerie médicale, Immunogénétique, Industrie, Industrie pharmaceutique, Infobiogen, Informatique, Inra, Inserm, Institut Cochin, Institut de myologie, Liaison génétique, Maladies génétiques, Maladies neuromusculaires, Maladies polygéniques, Maladies rares, Marqueurs, Médecine, Médecine prédictive, Microsatellites, Millennium, Ministère de la recherche, Ministre de la santé, Mucoviscidose, Mutations, Myopathies, Nanisme, Neurosciences, Obésité, OGM, Organismes modèles, Parthénogenèse, Patrimoine génétique, Pédiatrie, Philosophie, Phylogénie, Physiopathologie, Plasmides, Politique de recherche, Politique industrielle, Politique scientifique, Polymorphisme, Programme génome, Programme génome humain, Propriété intellectuelle, Protéomique, Pucés, Recherche clinique, Recherche et développement, Recherche génétique, Réparation de l'ADN, Risque génétique, Santé, Séquençage, Sociologie, Sondes, Statistiques, Technologie médicale, Téléthon, Tests génétiques, The Sanger Centre, Thérapie cellulaire, Thérapie génique, Transfert d'ADN, Transgène, Transgénose, Transsexualisme, Très grand séquençage (TGS), Trisomie, Unesco, Université, Vecteurs, Vieillesse, Virus, Workshop, Yacs,)

Par son **Origine** (service verseur, titre de la publication, etc.)

Par sa **Nature** (archives, rapport, publication, etc.)

Par le ou les nom(s) **d'auteur(e-s)**, le cas échéant

A chaque notice est rattaché un lien hypertexte **Document** renvoyant à la pièce dématérialisée sous forme de fichiers pdf, doc, jpeg, gif, jpg, docx, rtf et ppt.

## MODALITES D'ACCES

L'accès à la base de données s'effectuera selon les modalités arrêtées et les conditions prévues par Genopole® qui en est propriétaire. Quant aux opérateurs appelés à intervenir dans son fonctionnement, ceux-ci sont tenus de respecter un engagement de confidentialité établi par ses soins le 12 mars 2013.